DOCUMENT RESUME

ED 408 524 CG 027 769

AUTHOR Forsyth, G. Alfred; And Others
TITLE Heuristics for Improving the Interpretation of Research Reports.
PUB DATE Aug 96
NOTE 29p.; Paper presented at the Annual Meeting of the American Psychological Association (104th, Toronto, Ontario, Canada, August 9-13, 1996). For related document, see CG 027 770.
PUB TYPE Information Analyses (070) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Students; *Data Interpretation; Evaluation Methods; Evaluation Needs; *Evaluation Research; *Evaluative Thinking; *Heuristics; Higher Education; *Interpretive Skills; Professional Education; Research Skills; Statistical Analysis; Thinking Skills
IDENTIFIERS *Vignettes

ABSTRACT

Many students view statistics as their worst college course. Four heuristics that can improve students' proficiency in statistics and in interpreting reports of research are presented in this paper. The heuristics guide students' judgments about significance, generalizability, cause-and-effect, and strength of independent-dependent variable relationships. Previous studies of students' abilities to interpret research indicated that few of them understood random sampling and random assignment research methodologies as determinants of generalizability and cause-and-effect conclusions. It is supposed that students may fail to interpret research correctly because of the exclusive attention given to factual knowledge and statistical procedures rather than to interpretation abilities. It is recommended that a large set of vignettes (research-report summaries) can provide the critical methods needed for assessing student interpretations. One way to improve students' interpretation-of-research abilities is to develop a taxonomy that guides students in their judgments. A set of taxonomies is provided that can further aid students in answering questions about vignettes. A sample interpretation-of-research vignette is presented in which seven questions are posed to students when interpreting any research report. The students' answers to the questions can identify specific interpretation problems. (RJM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 408 524

HEURISTICS for IMPROVING the INTERPRETATION of RESEARCH REPORTS

G. Alfred Forsyth
(Millersville University)

Peter H. Bohling
(Bloomsburg University)

Richard B. May
(University of Victoria)

Presented at the 1996 Annual Meetings
of the American Psychological Association

G. Alfred Forsyth
Department of Psychology
Millersville University
Millersville, PA 17551

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A. Forsyth

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CG027769

BEST COPY AVAILABLE

ABSTRACT

We present four heuristics that improve students' proficiency in interpreting reports of research. These heuristics guide students' judgments about: significance, generalizability, cause-and-effect, and strength of independent-dependent variable relationships. They help overcome frequently made interpretation errors.

The heuristics are presented within the context of addressing questions to be answered in reading any report of research findings. We assess proficiency by having students interpret research report vignettes designed within the systematic framework for studying the understanding of interpretation of research concepts (Forsyth, Bohling, & Altermatt, 1995). The use of the heuristics is presented with a sample vignette.

Heuristics for Improving the Interpretation of Research Reports

In summarizing the outcome of a statistics education conference sponsored by the American Statistical Association, Hogg (1991) indicated that many students view statistics as their worst college course. He suggested that statistics courses

should have a greater focus on having students ask and answer research questions and interpret research findings than on using formulae to perform statistical calculations. The American Association for the Advancement of Science emphasized the importance of interpretation skills in their 1993 Benchmarks for Science Literacy. Similarly, a 1992 National Science Foundation report suggested that the development of abilities to apply knowledge about statistics has not kept pace with either rote-memory or calculation knowledge of statistics.

Forsyth, Bohling, and May (1991) used research report vignettes to assess several aspects of students' abilities to interpret research. Specifically, we administered an interpretation assessment instrument to students in first-and-second-level statistics classes in economics, mathematics, and psychology at three universities. This study explored random sampling and random assignment research methodologies as determinants of generalizability and cause-and-effect conclusions. Across the three disciplines at the three universities, independent of statistics-course level, there was little evidence that these relationships were understood. That is, generalizability ratings were not significantly greater for random-sampling than for available-group research reports. Similarly, confidence in drawing cause-and-effect conclusions was not significantly greater for random-assignment research reports than for

classificatory-independent-variable research reports.

Students were directed to interpret the research-report vignettes based on the descriptions of the research methods used in each study. Despite these directions, judgments about generalizability and cause-and-effect were based on the availability heuristic. That is, students' judgments significantly correlated with their beliefs about the existence of the independent/dependent variable relationship. If students believed that the independent and dependent variables were related, they expressed confidence in both generalizability and cause-and-effect. If students believed that the variables were not related, those students were not confident in generalizing the results or in drawing cause-and-effect conclusions. We refer to this over-reliance on one's initial belief as the availability heuristic. An availability heuristic error occurs when reliance on life experiences leads to a different interpretation than if judgments were based on research-methods information. May and Hunter (1988) report similar interpretation-of-research errors related to random sampling and random assignment.

Several other interpretation-of-research errors have been identified. Rosnow and Rosenthal (1989) examined errors in making judgments about the strength of a reported independent-dependent variable relationship. They noted an inappropriate reliance on reported p-value as an index of strength of effect.

Forsyth, Bohling, and May (1991) found that students inappropriately learned in their statistics courses to base their research-report interpretations on statistical formulae used to analyze the data rather than on design and research-methods information. For example, students were less confident in making cause-and-effect judgments if a Pearson r were used to analyze the data from a two-group study than if a t-test were used to analyze the same data.

We propose that one of the major reasons for students failing to interpret reports of research correctly is the exclusive attention to factual knowledge and statistical procedures rather than interpretation abilities in course examinations. As Garfield (1992) points out, students learn to value what they know know will be assessed. If a statistics teacher presents statistics merely as quantitative descriptions (e.g., the man is 5'11" tall; 70% of the freshman class is female), students in that course will not develop interpretation skills. If a teacher presents statistics merely as statistical significance testing in order to make inferences (e.g., children read to regularly have significantly higher language arts scores than children not read to), students in that course will not develop interpretation skills. If a teacher presents statistics merely as the computation of probabilities (e.g., how much more likely is one to obtain two sixes if she rolls five dice than if she rolls two

dice?), students may become more proficient at Yahtzee, but will not develop interpretation skills. If a teacher presents statistics merely as using a computer to obtain means, variances, standard deviations, Z-scores, correlation coefficients, and t-tests, students will not develop interpretation skills. If statistics teachers have the development of interpretation skills as a course goal, they must provide students with interpretation-of-research exercises and must assess those skills.

A major reason for faculty not assessing students' abilities to interpret reports of research is that reading a complete journal article is time intensive. Students would have to read several jargon-laden, lengthy, and perhaps boring and trivial articles in order for a teacher to assess their abilities. Another consideration is the challenge for faculty to find research articles that differ systematically in features such as random sampling vs available groups, random vs classificatory assignment of subjects, number of subjects per group, p-value, and levels of strength-of-relationship indices. Media reports of research do not provide a viable alternative because they usually do not contain sufficient information for the reader to draw appropriate conclusions. Without a readily available assessment instrument, faculty tend not to assess interpretation skill development.

One solution to this problem is the development of a large

set of vignettes (research-report summaries) that provide the critical methods and results information needed for interpretation. Features such as those in the previous paragraph (e.g., presence or absence of random sampling) can be varied across vignettes. After reading each vignette, students would answer a set of questions to assess their interpretation skills. Forsyth, Bohling, and Altermatt (1995) presented such an assessment instrument along with guidelines for its use.

In addition to designing ways to assess students' interpretation-of-research abilities, we are interested in developing strategies for the improvement of those abilities. We first developed a teaching strategy to guide students in their judgments about internal and external validity. This consisted of a two-by-two taxonomy for categorizing research studies in terms of random sampling vs available groups and random vs classificatory assignment of subjects to levels of the independent variable. This taxonomy and its use in judging internal and external validity was presented by Forsyth and Bohling (1994) and by May, Masson, and Hunter (1990).

To test the effectiveness of instruction using this taxonomy, Forsyth, Arpey, and Stratton-Hess (1992) randomly assigned students to either a taxonomy study condition or a control study condition. Instructional materials in the control condition consisted of quotations from 10 current research methods and

statistics textbooks. These quotations were clustered under topic headings appropriate to the questions in the interpretation assessment instrument. Both random sampling and random assignment were among the topic headings. The taxonomy study condition subjects were trained to use the two-by-two taxonomy to classify studies based on the presence or absence of random sampling and random assignment. The analysis of data from that study indicated that students instructed in the use of the taxonomy made appropriate judgements about generalizability and cause-and-effect. The control subjects, using the textbook quotations, relied on the availability heuristic in making their generalization and cause-and-effect judgements and therefore made significantly more errors.

The purpose of the present paper is to introduce additional taxonomies that can be used as heuristics to assist students in answering questions about reports of research. Using a sample research-report vignette, we first suggest seven questions to be answered in the process of reading a research report. We then introduce heuristics that we have used to guide students in answering four of these seven interpretation-of-research questions.

Questions to Guide the Interpretation of Research

Forsyth, Bohling, and Altermatt (1995) proposed the use of research-report vignettes along with seven questions designed to assess students' interpretation-of-research abilities. They also provided a large set of vignettes and a systematic framework for altering vignettes to assess the understanding of specific interpretation-of-research concepts. The following is a sample vignette with related interpretation-of-research questions.

Sample Interpretation-of-Research Vignette

School counselors Rhonda Flaboff and Wanda B. Heer were interested in the relationship between school absences and being involved in an after-school fitness program. Using student records at the high school where they served as counselors, Rhonda and Wanda identified 1,637 students who were not on an athletic team nor involved in any other regularly scheduled after-school activity. These students were sent questionnaires asking them to indicate their interest in 10 after-school clubs. The fifth club was "Fun Fitness" with activities such as doubles tennis, swimming, volleyball, and recreational soccer. Four hundred fifty students indicated an interest in the fitness club.

Rhonda and Wanda randomly sampled 64 of these 450 students for the study and randomly assigned each of these 64 students to either the fun-fitness group or the control group. Each group consisted of 16 females and 16 males. The fitness club met after

school three times a week for five months. Physical education teachers Dennis Anywon and Val E. Bahi coordinated club activities that provided physical fitness exercise in an enjoyable social setting. The control group was called to a January meeting and asked to rank-order a list of fitness activities so that Dennis and Val could purchase needed materials for club activities to which the control-group students would be invited the following Fall.

Rhonda and Wanda kept a record of the number of school absences during a five month period for each of the 64 students in the study. Mathematics teachers, Olive Nombers and Cal Q. Lator, analyzed these data and reported that the fitness group had fewer school absences than the control group, $p<.05$. Olive said that the fitness vs no-fitness variable accounted for 60% of the variation in school absences. The 95% confidence interval for the difference in mean absences predicted that the interested-in-fitness population would average between two and 10 more absences if that population were not in a fitness program than if they were in the fitness program.

Based on the research methods used and the reported results, please answer each of the following questions:

1. **What are the independent and dependent variables in this study?**

   independent variable is _____.
   dependent variable is _____.

2. How confident are you that the results indicate that there is a relationship between the independent and dependent variables? (circle your answer)

1	2	3	4	5
not very confident		moderately confident		very confident

3. How confident are you that the results of this study could be <u>generalized</u> to the other high school students interested in the Fun-Fitness Club? (circle your answer)

1	2	3	4	5
not very confident		moderately confident		very confident

4. How confident are you that participating vs not participating in the fitness club activities <u>caused</u> the difference between the group means?

1	2	3	4	5
not very confident		moderately confident		very confident

5. How strong do you consider the relationship between the independent and dependent variables to be? (circle your answer)

1	2	3	4	5
not very strong		moderately strong		very strong

6. How important do you consider the finding about the relationship between variables to be? That is, how important is it to get uninvolved high school students to participate in a fitness program? (circle your answer)

1	2	3	4	5
not very important		moderately important		very important

7. Is there critical information missing in this report that is needed for interpretation?

YES _____ NO _____

If yes, what information is needed?

Expressed in general terms, we recommend that the following seven questions be answered when interpreting any research report:

1. What are the independent and dependent variables?
2. Was a relationship found between the independent and dependent variables?
3. To what extent can the results of the study be generalized? That is, how appropriate is it to infer that the independent-dependent variable relationship would also exist for others than those in the study?
4. How appropriate is a cause-and-effect conclusion? That is, did changes in the independent variable cause changes in the dependent variable?
5. How strong is the relationship between the independent and dependent variables?
6. How important do you consider this finding about the independent and dependent variable relationship to be?
7. What additional information should have been provided to permit a clearer interpretation of the research?

Heuristics

This vignette-with-questions assessment procedure not only identifies specific interpretation problems, but can also lead to the development of teaching strategies for improving

interpretation skills. In this section, we introduce taxonomies to help students correctly answer questions 2, 3, 4, and 5.

Because of the reference to independent and dependent variables across questions and frequent confusion of the independent and dependent variables by students, we recommend that students first identify the independent and dependent variables in the study. To guide the question 1 process, we recommend that Figure 1 be used as a student worksheet. For each study, students would be asked to use the right-most block to indicate how the researcher operationally defined the behavior of interest (the dependent variable). In the sample vignette, this is the number of school absences. Students would then be asked to write in the top left block what independent variable was examined in the study. In the sample vignette, this is participation or no participation in the fun-fitness program. To help students understand that other variables could also account for variability in the behavior of interest, they would be asked to propose other possible independent variables that could account for variability in the behavior of interest. Examples of such variables within the context of the sample vignette are: 1) medical health of the student, 2) amount of alcohol consumed, and 3) number of cigarettes smoked per day. For each extraneous variable, students would be asked how they would determine if it is confounded with the independent variable in this study and how

they would carry out the study to prevent confounding.

To help students understand that other variables may statistically interact with the independent variable, they would also be asked to propose one such variable. Within the context of the vignette, an example of a potentially interacting variable would be the number of close friends attending the same school as the participant. It may be that the fun-fitness club will reduce absences for those with no close friends in the school, but have no effect on school absences for participants with many close-friend school mates.

Insert Figure 1

The second question to be answered is whether or not a study should be considered as having found a relationship between the independent and dependent variables. This requires that the students have an understanding of what a null distribution is and the meaning of statistical significance. Figure 2 presents a taxonomy to guide students in their choice of a criterion for statistical significance. Within the context of this taxonomy, how rigorous a criterion is used would depend upon the cost of a Type I error and whether the study is the sole basis for decision

making or is part of a meta analysis. Within the context of the sample vignette, if participation in the fun-fitness program has no effect on school absences for the population of interest, a decision to reject the null hypothesis would constitute a Type I error. The cost of that error would be high if the decision to reject the null hypothesis led to the reader instituting an extensive fun-fitness program in his/her school district. The cost of a Type I error would be lower if the decision to reject the null hypothesis would simply mean that a small trial program would be carried out in the reader's school district. As Figure 2 indicates, the alpha level would be smaller when the Type I error cost is high. Thus, if decision making is based on a single study, a high-cost Type I error might result in setting alpha at .001 while a low-cost Type I error might lead to using an alpha of .05.

A second factor that influences the criterion for rejecting the null hypothesis is whether the reader will take action based solely on one study or whether this study is one of several the reader will examine prior to taking action. Thus, if a meta analysis is to be carried out to decide whether or not to institute a district-wide,fun-fitness program (high cost), the criterion for considering this study as supporting that decision might shift from an alpha of .001 to an alpha of.05.

Insert Figure 2

The third question asks about the degree to which the results generalize to others than those in the study. That is, students are asked to judge the external validity of a study. As indicated in Figure 3, studies in which participants are selected randomly from some defined population have a higher external validity than studies using an available group. In the sample vignette, 64 participants were randomly sampled from a population of 450 students interested in the fun-fitness club. Thus, it would be appropriate to infer that there would be fewer absences for the population of 450 students if they all participated in the fitness program. If Rhonda and Wanda had simply used an available group, the external validity would have been low. Replication is another factor to be considered in judging external validity. If a study is presented as a replication of a previous study using different participants and the results of the replication study are consistent with the initial study, there should be increased confidence that the results generalize to others than those in either study alone. Suppose Rhonda and Wanda indicated that their results were in accord with those of another study examining the effect of fitness-program

participation on school absences. This would increase the reader's confidence about the results generalizing to others than those in Rhonda's and Wanda's study.

Insert Figure 3

Question four asks students to indicate their confidence in concluding that changes in the independent variable caused changes in the dependent variable. Random assignment of participants to experimental groups is an important research method for increasing internal validity. Rhonda's and Wanda's confidence that the fitness program caused a reduction in school absences was increased by their randomly assigning participants to the two groups. If they had simply identified students who were or were not participating in a fun-fitness program, their confidence in drawing cause-and-effect conclusions would be lower. When subjects assign themselves to levels of the independent variable, the independent variable is said to be classificatory. Internal validity is lower in such studies because other variables may be confounded with the independent variable. If the fitness variable were classificatory, confounding with extraneous variables would be more likely. For

example, more participants in the no-fitness group than in the fitness group may have been smokers. The smoking rather than fitness participation might have caused some or all of the differences between groups. To ensure that groups are alike on a specific extraneous variables that might be related to the dependent variable, researchers might: (1) hold that variable constant, (2) match participants on the extraneous variable when creating the groups, or (3) cross the independent variable with levels of the extraneous variable. These procedures eliminate that extraneous variable as explaining why the groups in the study are different. In Rhonda's and Wanda's study, the groups could be made equivalent in terms of amount of smoking by assigning participants to groups so that the mean amount of smoking is the same for both groups. Alternately, within each level of the fitness independent variable, equal numbers of nonsmokers, half pack per day, pack per day and one and one-half pack per day participants could be selected for the study.

Insert Figure 4

The fifth question asks students to judge the strength of the relationship between the independent and dependent variables.

Figure 5 presents a taxonomy to guide the search for information to make this judgement. In studies such as the sample vignette, the readers have a life-experience-based concrete referent for the dependent variable. That is, they can relate to high school students missing 0, 5, 10, or 15 days of school over a five month period. In such cases, the use of a confidence interval as a measure of strength of effect tells the readers how many fewer absences the average individual in the population would have if he/she participated in the fitness program. Suppose the dependent-variable construct were social self-esteem and it had been measured with an unpublished social-self-esteem scale. A 95% confidence interval might indicate that the population would have a social-self-esteem average between two and 10 points higher if all population members were in the fitness rather than the no-fitness condition. Without any concrete understanding of what two or 10 means, the confidence interval conveys little other than that statistical significance at $p<.05$ was found. When the dependent variable has no concrete referent, readers must rely on eta-squared (r-squared if linear) to judge the strength of effect.

When the major purpose of a study is to compare the relative strength of two or more independent variables, eta-squared (r-squared if linear) provides an index of the relative success of each independent variable. When the dependent variable in

multiple-independent-variable studies has a concrete referent (e.g., school absences, grade point average), eta-squared (r-squared if linear) indicates the relative success of an independent variable and the confidence interval indicates how much change takes place in the dependent variable as a function of change in the independent variable.

The sixth question asks subjects to indicate how important they consider the independent-dependent relationship to be. This involves a subjective judgment without any single correct answer. In discussing importance judgments, students should be asked to analyze the degree to which each of three factors played a role in their judgment. Students may judge the finding to be important to them personally, important to society, or important to science.

The seventh question asks subjects to identify any additional information that they need to interpret the research report. For example, in the sample vignette, it would be important to know if the nonfitness group had a much larger standard deviation than the fitness group. It may also help to know how many students from each group participated in a fitness program outside the school setting.

We have found the heuristics presented in this paper to be useful in improving our students' interpretation-of-research abilities. What is needed next are formal assessments of the

effectiveness of these heuristics. Those investigations are currently underway. We invite you to increase the generalizability of findings by assessing your students' interpretation-of-research abilities with our vignettes and taxonomy heuristics.

Insert Figure 5 here

References

American Association for the Advancement of Science (1993). Benchmarks for science literacy, Oxford University Press.

American Association for the Advancement of Science (1989). Science for all Americans. Washington, D.C.

American Association for the Advancement of Science (1992). Indicators of science and mathematics education. Washington, D.C.

Forsyth, G.A., Arpey, S.H. and Stratton-Hess, C.L. (1992). Correcting errors in the interpretation of research. Paper presented at the 100th Annual Meeting of the American Psychological Association, Washington, D.C.

Forsyth, G.A. and Bohling, P.H. (1994). Interpreting statistics and behavioral research. Acton, Mass.: Copley Publishing Company.

Forsyth, G.A., Bohling, P.H. & Altermatt, T.W. (1995). Developing and assessing students abilities to interpret research. Paper presented at the annual meetings of the American Psychological Association, New York, N.Y,

Forsyth, G.A., Bohling, P.H., and May, R.B. (1991). Factors affecting the misinterpretation of research. Paper presented at the 99th Annual Meeting of the American Psychological Association.

Psychological Association, San Francisco, California.

Garfield, J. (1992). How students learn statistics. Presented at the American Statistical Association Winter Meeting. Louisville, Kentucky.

Hogg, R. V. (1991). Statistical education: Improvements are badly needed. The American Statistician, 45, 342-343.

May, R. B. and Hunter, M. A. (1988). Interpreting students' interpretations of research. Teaching of Psychology, 15, 156-158.

May, R. B., Masson, M. E. J., and Hunter, M. A. (1990). Application of statistics in behavioral research. New York: Harper and Row.

National Science Foundation (1992). Indicators of science and mathematics education. Washington, D.C.

Rosnow, R. L. and Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. American Psychologist, 44 (10), 1276-1284.

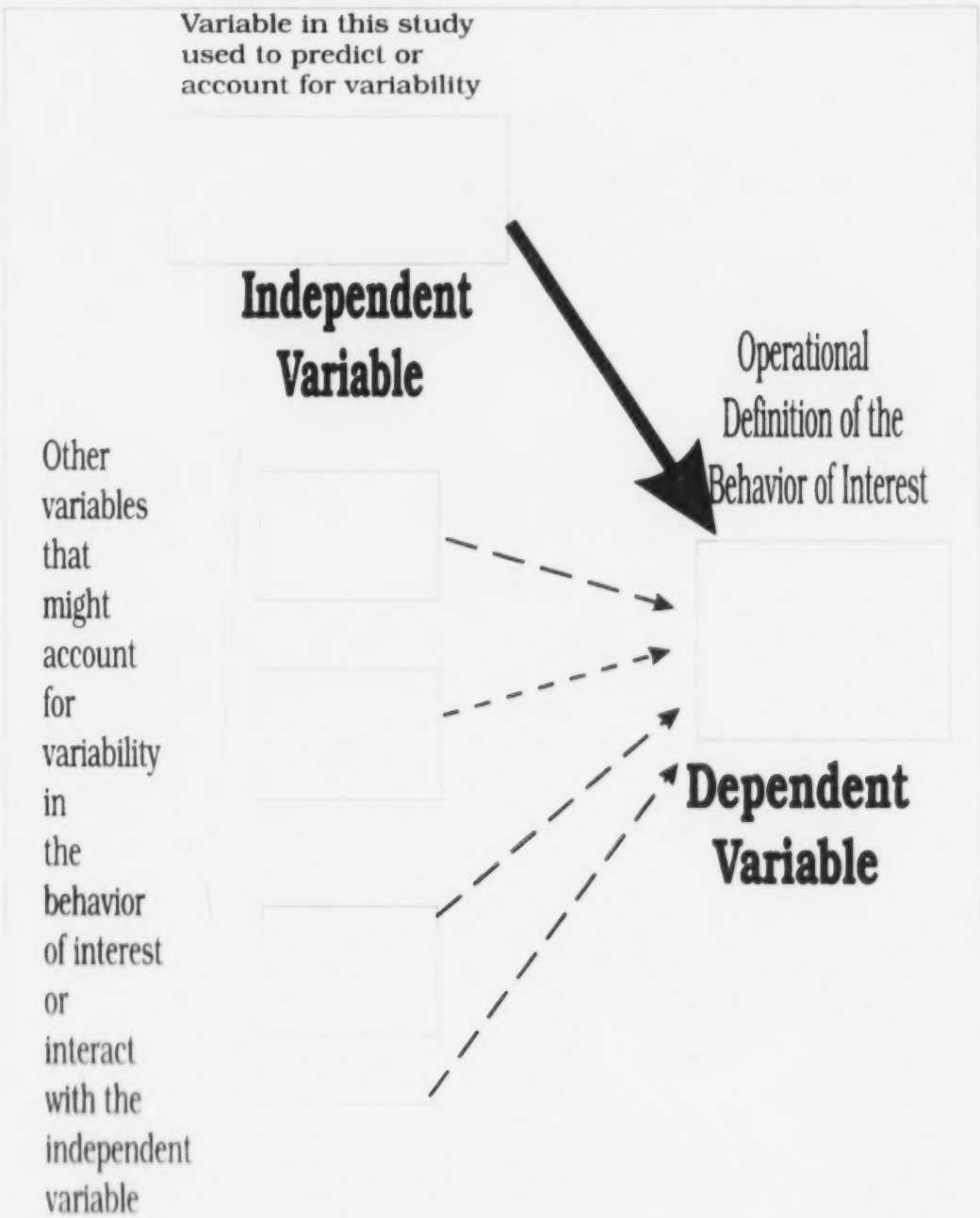
Variable in this study used to predict or account for variability
Independent Variable
Operational Definition of the Behavior of Interest
Other variables that might account for variability in the behavior of interest or interact with the independent variable
Dependent Variable

Figure 1
Student Worksheet for Identifying Independent, Dependent, and Extraneous Variables

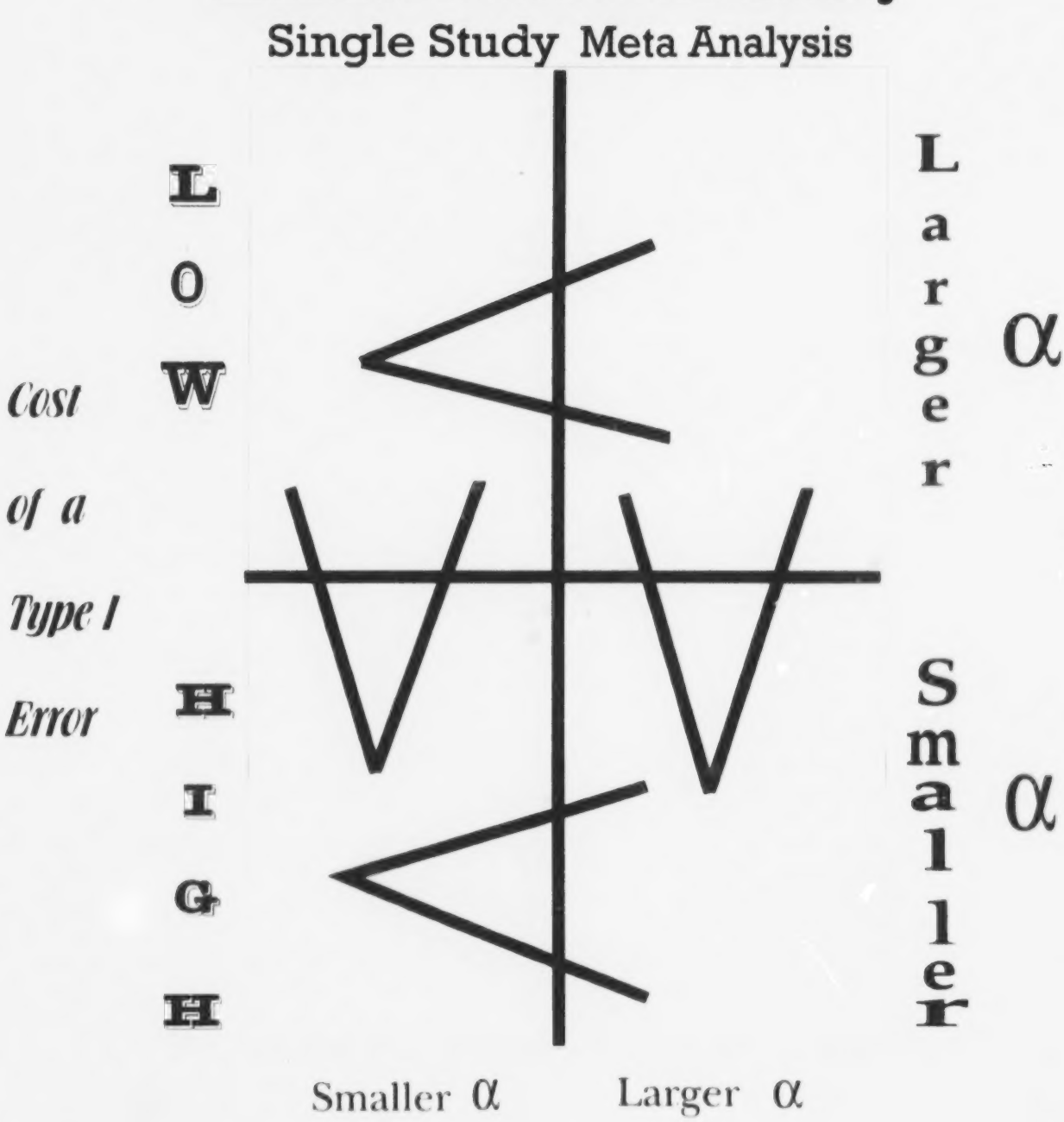

Figure 2
Taxonomy for Setting the Statistical Significance Criterion (alpha level)

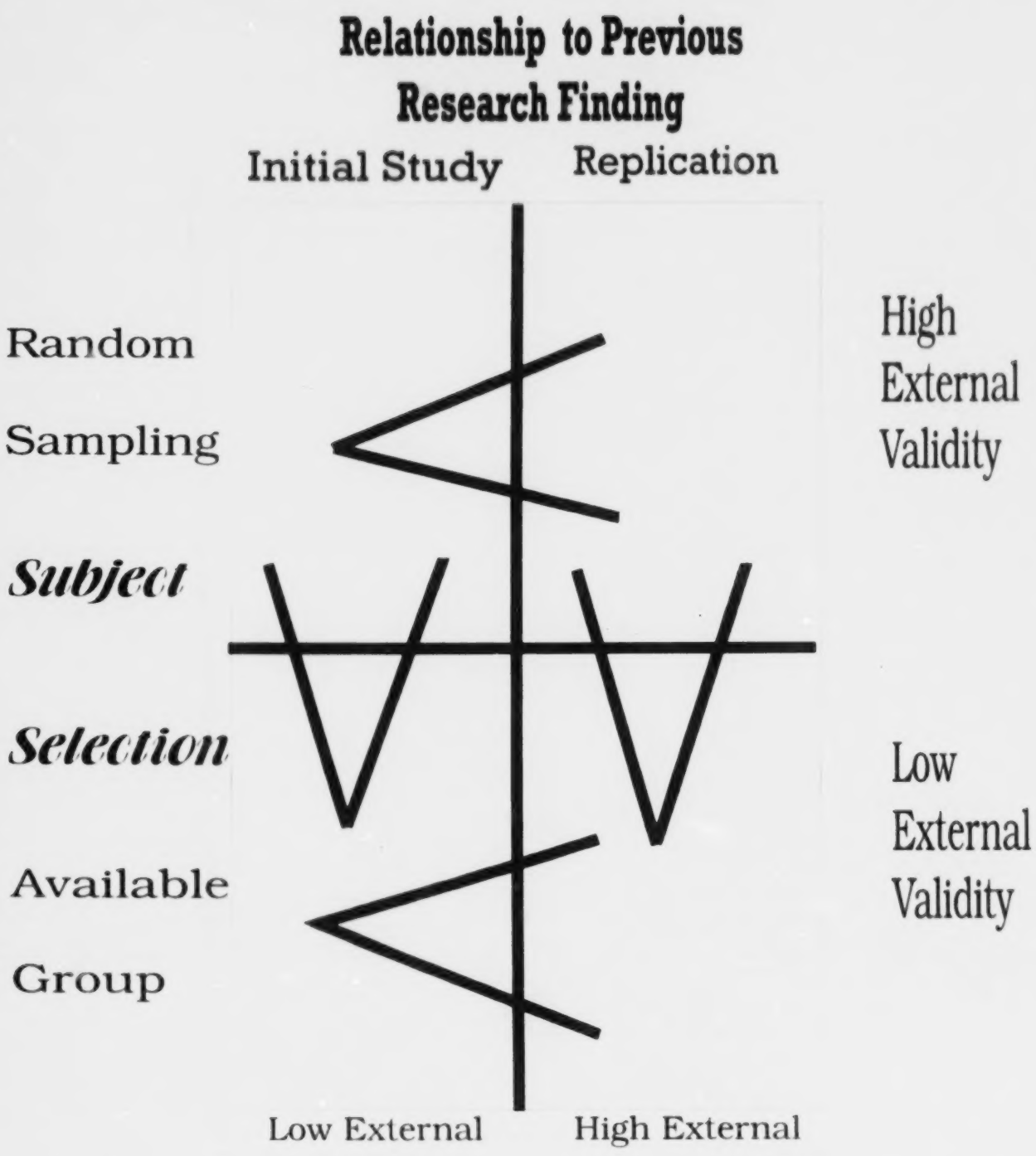

Figure 3
Taxonomy for Judging the External Validity of a Study (generalizability)

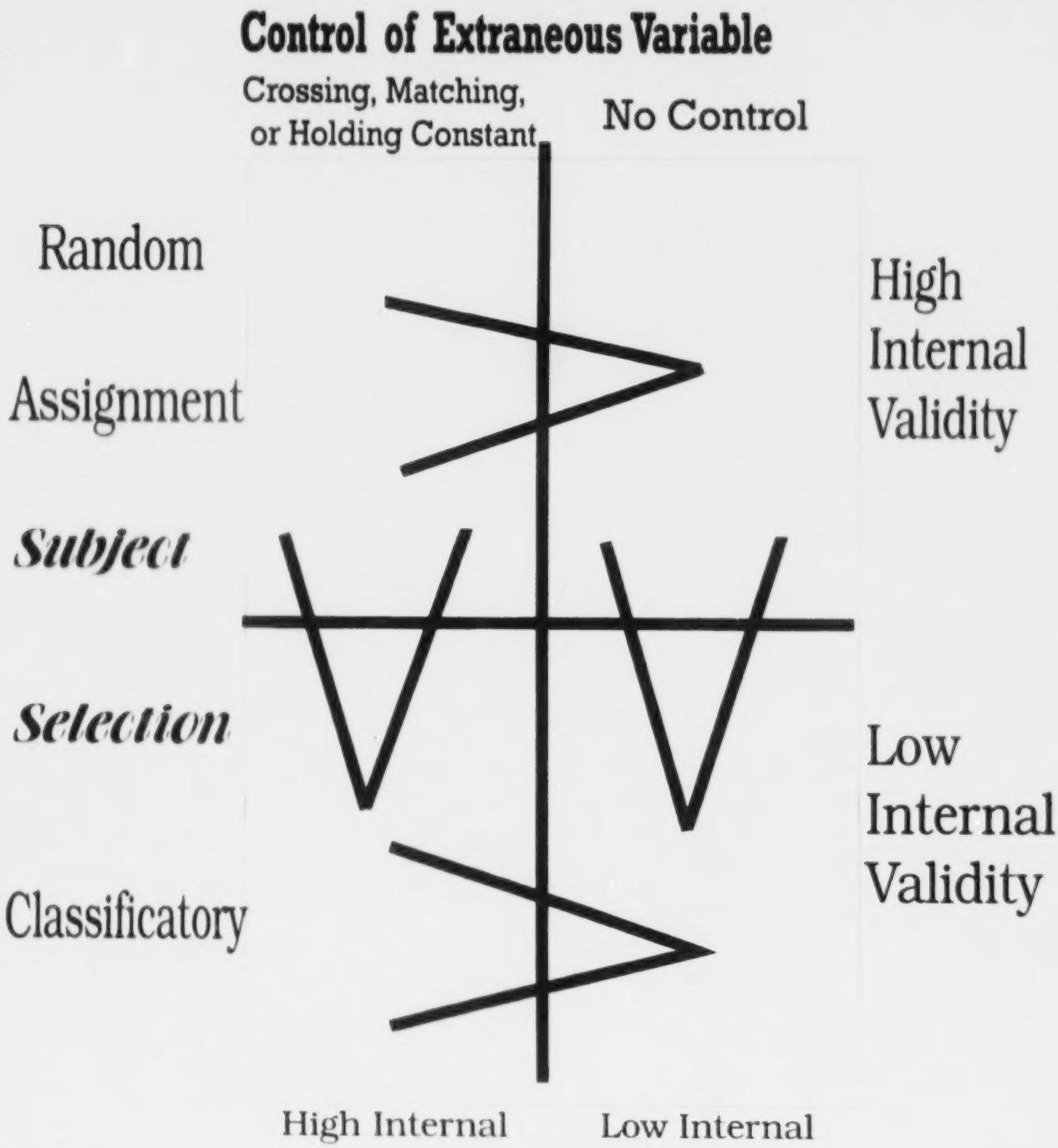

Figure 4
Taxonomy for Judging the Internal Validity of a Study (cause and effect)

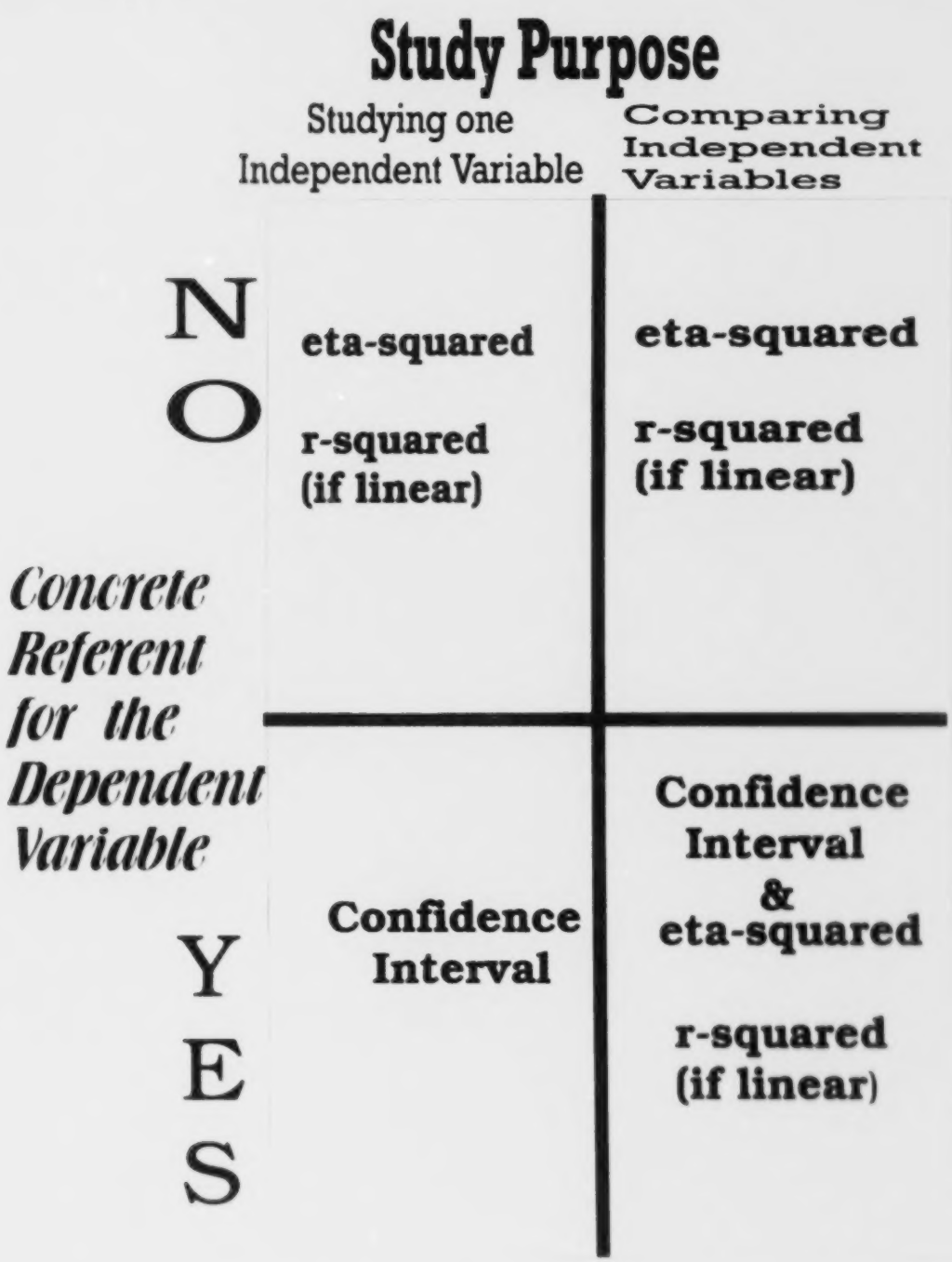

Figure 5
Taxonomy for Judging the Strength of the Independent-Dependent Variable Relationship